WATER QULITY SENSING USING MULTI-LAYER PERCEPTRON ARTIFICIAL NEURAL NETWORKS

Ayman F. Batisha[1]

ABSTRACT: The purpose of this paper is to evaluate the neural network approach in the context of operational water quality sensing. The classification of water quality data is a typical pattern recognition problem that poses many difficulties. Traditional methods for classifying high volumes of such data into large numbers of classes based on statistical parametric methods often do not give sufficient descriptive accuracy for discriminating the numbers of classes required. The use of multilayer pcrceptron neural networks as a new method for solving this problem for realistic operational purposes has been established. The multilayer perceptron offers a good classification method and competes well with the traditional techniques used in statistical parametric methods. Indeed by using reasonably large network architectures, the method seems to work quite well with large numbers of classes that is where problems are normally encountered with the traditional parametric methods. The neural networks have much potential in water quality sensing and I hope to integrate them into operational applications in the future.

1. INTRODUCTION

The classification of water quality sensing data is a typical pattern recognition problem that poses many difficulties. High-resolution images obtained from Earth observation water quality data such as BOD and COD have many potential uses, particularly in mapping the environmental setting on the ground surface and in monitoring Quality of Life. For operational purposes the pollutant water images usually need to be classified into a large number of classes representing the diversity of surface water cover. Traditional methods for classifying high volumes of such data into large numbers of classes (>20) based on statistical parametric methods often do not give sufficient descriptive accuracy for discriminating the numbers of classes required. We have been investigating the use of multilayer perceptron networks as a new method for solving this problem for realistic operational purposes.

Neural networks are among current artificial intelligence (AI) research areas of interest to the environment industry. Neural networks attempt to model the brain learning, thinking, storage, and retrieval of information as well as associative recognition. Since the early 1950s and despite their start research in neural networks has grown slower systems due to initial skepticism. By the mid 1980s neural networks research started to flourish, and large number of expert systems emerged in a diversity of fields.

Unlike these AI-based systems, traditional decision-analysis techniques (probabilistic methods, multiattribute utility theory, fuzzy sets theory, etc.) generally require advanced mathematics, making them less acceptable for practicing environmental personnel.

---

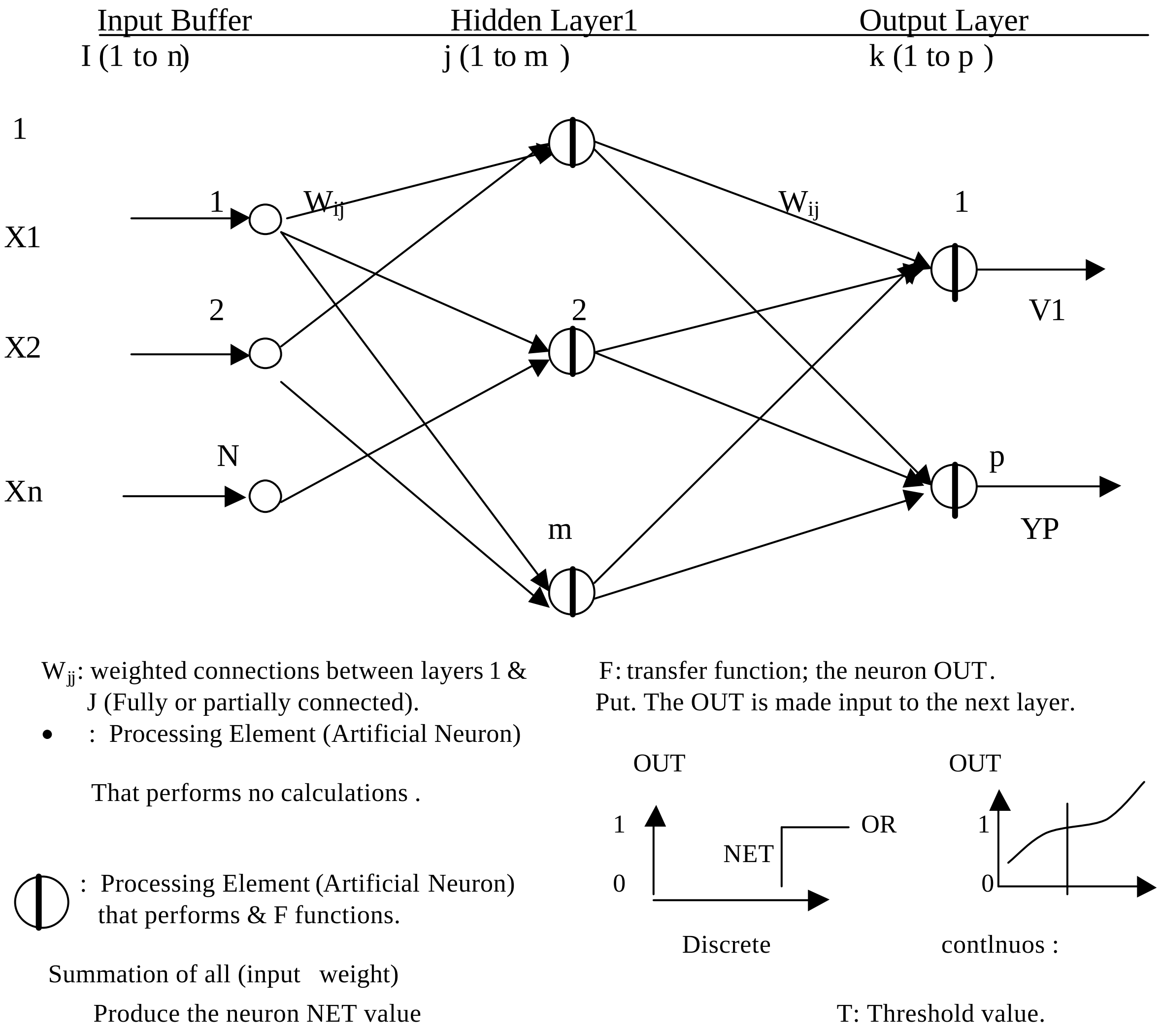
[1] *Researcher, Environment and Climate Research Institute, National Water Research Center, Cairo, Egypt*

The objective of this paper is to introduce networks (NN) as promising management tools that have capabilities particularly suited for analogy based decision problems, so prevalent in all levels of environmental engineering and management tasks. The parallel and distributed structure of neural networks along with their capabilities of generalization, fault tolerance, adaptive and associative performance, ability to perform dynamic and real-time functions, and their limited requirement of software, ensure their appropriateness for many practical environmental applications. A procedure for casting environmental problems into patterns is presented and the variables that govern the design of a neural network for a particular problem are discussed. The different characteristics of commonly used neural network architectures are described in a simple manner irrespective of their sophisticated mathematical formulations. An example neural network for optimum markup estimation is presented.

While expert systems are procedural software systems that attempt to mimic the problem-solving function of the brain, neural networks are introduced as hardware or software systems analogous to biological neural systems both in structure and in functionality. Despite the diversity of network paradigms (architectures), they consist of similar components. In the general from of a neural network, the unit analogous to the biological neuron is referred to as processing element (PE). The network consists of those elements usually organized into a sequence of layers or slabs with full or partial connections between successive layers specifically designated. Fig. 1 shows a simple two-layer network architecture. The neural network has an input buffer (not considered as a layer to which to data is presented to the network, and an output layer which holds the response of the network to a given input. Layers distinct from the input buffer and the output layer are called hidden layers. As shown, in Fig 1, a processing element (artificial neuron), usually excluding those in the input buffer, performs summation (Σ) and transfer function (F) to determine the value of its output. in essence, a set of inputs, each representing the output of another neuron in a preceding layer, are received by the processing element (PE). Each input is multiplied by its connection weight (analogous to the human synaptic strength) and all the weighted inputs are then summed to produce the neuron “NET” or activation value, which is then modified.

The transfer function can be discrete threshold function which passes information if the value of the NET reaches a certain level [i.e., PE output is one, if NET > T (threshold value); otherwise, the output is kept zero, or it can be a continuous function of the NET value (output is a real number from 0 to 1.0). neural networks, generally, could be of two types: (1) “Feed-forward” or nonrecrrent, where the network PE connections and thus the information flow are in one direction as shown in Fig. 1; “recurrent,” which exhibits a more general network structure that allows feedback connections through weights extending from one layer to another or to itself.

There are two main phases in the operation of a neural network: learning and recall. Learning is the process of adapting the connection weights in response to a number of examples (stimuli) being presented at the input buffer and, optionally, at the output buffer. The task is to arrive at unique set of weights that are capable of correctly all example pattern (s), used in learning, with their desired output pattern(s). There are two main types of learning: Supervised and unsupervised. Supervised learning refers to the case in which the network is presented with some input examples and their desired responses (association). The desired outputs are used in this case to teach the correct responses. Unsupervised learning on the other hand, refers to the case when only input examples are presented to the network, and the network must organize itself on the basis of the input stimuli it receives.



W_{jj}: weighted connections between layers 1 & J (Fully or partially connected).

- : Processing Element (Artificial Neuron) That performs no calculations .

: Processing Element (Artificial Neuron) that performs & F functions.

Summation of all (input weight) Produce the neuron NET value

F: transfer function; the neuron OUT. Put. The OUT is made input to the next layer.

T: Threshold value.

FIG. 1. Example of Simple Neural Network

Learning session, learning examples are shown to the network for many thousands of times until a certain preset to stop the learning session is met. One such criterion is to consider the network to have adequately learned when the error between the output predicted by the network and the desired output, accumulated in all learning examples, is less than a specified limit. Learning sessions often consume large amounts of computer time and can face serious problems including network "paralysis" and "local minima." Network paralysis refers the case when the weights unpredictably come to a standstill status and adjust during training. Local minima, on the other hand, refer to the case when the weights settle on a less than optimum status. Training algorithms also may include deterministic or statistical procedures for the network weights adjustment. Statistical procedures have been used to alleviate the local minima problem and when high network accuracy is desired despite the longer training time it requires. If the network training is successful, the network is expected to produce outputs that are satisfactorily close to the desired set of output(s) used in training.

Neural network enjoy several characteristics that distinguish them from other AI traditional architectures including expert systems, some of these characteristics include:

1. They are particularly suited for pattern recognition tasks where large number of attributes must be considered in parallel.
2. They learn by example, Unlike expert systems, neural networks learn many example patterns and their associations, i. e., desired outputs or conclusions. These examples could be elicited from experts without the need for asking how and why they cane to those conclusions. Problems generally associated with knowledge acquisition are therefore eliminated.
3. They produce fast responses, irrespective of their requirement of large computer time for learning. This is due to the parallel structure of neural networks.
4. They could extract classification (clustering) characteristics from a large number of input examples, as in the case of unsupervised learning. If, for example, a large number of field data are collected from a construction site, a suitable network can identify the different clusters (groups or classes) that characterize the whole population.
5. They have distributed memory; the connection weights are the memory units of the network. The value of the weights represent the current state of knowledge of the network. A unit of knowledge. Represented for example by an input/desired output pair, is distributed across all weighted connections of the network.
6. They have associative memory. The network responds an accretive or interpolative way to noisy, incompetent, or previously unseen data. An auto associative network, where input is

equal to desired output, can produce a full output if presented with a partial input. This property is called “generalization.

7. They are fault-tolerant. Since memory is distributed, failure of some pro-cessing elements will slightly alter the overall behavior of the network. However, failure of any samall part in a traditional computing system will stop its performance. The characteristic is very well suited to applications where reliable system need to be developed from less reliable components.
8. They could represent uncertainty; a measure of “belief” could be incor-porated by modifying the problem pattern in ways: (!) by selecting input values to represent a measure of belief in the attribute: and (2) by adding another attribute representing the measure of belief in the input example.
9. They require a lesser amount of storage memory. Since there is only one set of network weights capable of representing a large space of stored patterns.

2. WATER QUALITY SENSING

The area of water quality sensing is continually being influenced by advances in science and engineering and by changing social and economic factors. The term water quality sensing refers to the overall health of surface water bodies including the interrelationships between the water and underlying sediments. In general, water quality sensing involves water quality monitoring and assessment, establishing achievable water quality goals and controlling pollutant loadings to reach and maintain desired water quality levels.

Water quality sensing monitoring and assessment begins with collecting and analyzing samples to measure various physical, chemical and biological parameters. The resultant data are then evaluated and compared with accepted criteria to develop an understanding of the health of the water body. These data can also be used as inputs to verify or calibrate water quality simulation models. Once proven to produce reliable results, these models can be used to predict and evaluate the water quality sensing benefits of varying levels of pollution control.

In establishing water quality sensing goals, the desired use of the water body is determined. Specific water quality levels, often referred to as water quality standards, are established to sustain such uses as providing safe drinking water, shellfish harvesting, safeguarding swimmers and bathers, and protecting aquatic life.

Pollution loads that cause the degradation of water quality sensing originate from both discrete and diffuse sources. Discrete sources, also known as point sources, include sewage treatment plants and wastewater discharges from industrial and mining facilities. Diffuse sources, often referred to as nonpoint sources, include contaminated runoff from agricultural, forestry, and urban areas. For the most part, discrete sources are easily identified and regulated. On the other hand, diffuse sources are difficult to isolate and control.

Pollution prevention measures that reduce the amount of pollutants generated and best management practices that attempt to alter land use to reduce pollution loads appear to be the most practical means to control diffuse sources. Technologies for the treatment of domestic and industrial wastewater are well established and documented. However, as treatment needs are increased to address water quality problems such as effluent toxicity, bioaccumulation, and sediment contamination, additional technologies must be developed. Controlling diffuse sources of water pollution is a more complex and challenging task. The use of conventional treatment techniques is limited due to the intermittent and widespread nature of these sources.

The foundation of sound water quality sensing management is reliable monitoring and assessment. Monitoring data result from physical, chemical and biological measurements of ambient waters, pollution sources, sediments, and biota.

3. WATER QUALITY SENSING SYSTEM FORMULATION

For testing purposes we have carried out some experiments on water cover mapping in a predominantly rural area of Lake Nasser. We have acquired multi spectral water quality data on two separate dates. The water quality data images have been spatially registered and each contains data in three spectral channels. Multi-temporal data is preferred for water cover mapping because seasonal changes in pollutants are known to be useful in discriminating certain water cover classes. In order to classify such imagery a training set of ground truth data was acquired from the test area by field survey.

The neural networks were able to produce highly accurate diagnoses. The samples used to train the neural networks can be developed either over time, as more pollution states are experienced by the real environmental system, or by carrying out a testing program that records the variations in monitoring signatures resulting from various pollution states introduced to a physical model of the environmental system. Both of these approaches are time-consuming. Therefore, analytical generation of training samples through mathematical modeling of the environmental system is a logical alternative. The changes in the state variable for monitoring corresponding to some of the pollution states were used to train two neural networks. The trained networks were then used to diagnose other pollution states. A typical sample case used to train the neural networks for pollution diagnosis would be composed of a correct pollution diagnosis and its corresponding changes in monitoring signatures. For each pollution case, the reading at sensing station is normalized with respect to the sensing station. The state variables used for monitoring are computed from

$$X_i = \frac{S_i - B_i}{B_i} \qquad [1]$$

Where

X_i = State variable for detection (input to the neural network)

S_i = Reading at sensing station i

B_i = Baseline value at sensing station i

The purpose of using the state variables defined in Equation (1) instead of the absolute values of the reading at sensing station to magnify the effect of the location of the pollution, where the pollution patterns assume their maximum values in the vicinity of the pollution location. For example, the pollution pattern corresponding to the first pollution states takes its maximum value at the first sensing station mounted on the first component. The neural network was trained to detect the pollution and to determine its class.

The system started with identifying a set of environmental signatures that can characterize the health condition of the environment. Then determining locations of sensing station based on knowledge of the sites of the weak links and the nature of the signatures used; and generating simulated pollution patterns by analytical modeling at the locations of the unsatisfactory links. Training properly designed neural networks with the analytically produced pollution patterns. In the system described in Figure 2, the neural networks will yields the best possible diagnosis of the incoming signal with respect to the pollution classes presented during the training session. Before starting a new diagnosis cycle, the signal/diagnosis data of the previous cycle is added to the training set. This step makes the system adaptive; i.e., the network will be able to detect the pollution source or class accurately if it occurs a second time. This will cause the network performance to improve and stabilize as it gets more pollution/diagnosis experience. The frequency of the cycle would depend on the importance of the environmental system and the stability conditions.

4. IMPLEMENTATION OF NEURAL NETWORK CLASSIFIER

The chosen neural model is a fully connected multilayer feed-forward network with symmetric sigmoid activation functions and trained by the back-propagation algorithm to minimize a quadratic error criterion. The input layer consists of 6 units yielding only the individual channel values, so that no context information is provided to the network. The output layer consists of 20 units mapping a local representation of the classes: for example, the target vector for a pixel belonging to class 3 would be (-1, -1, 1, -1, -1...-1). The network's response to a given input is determined by the output unit having the Highest Activation State, even if that maximal value is negative. Although this is the weakest possible criterion it offers the advantage of providing the following qualitative measure of confidence level: the average value of maximal firing on the output units was substantially high. The correctly classified unit than for mis-classification cases.

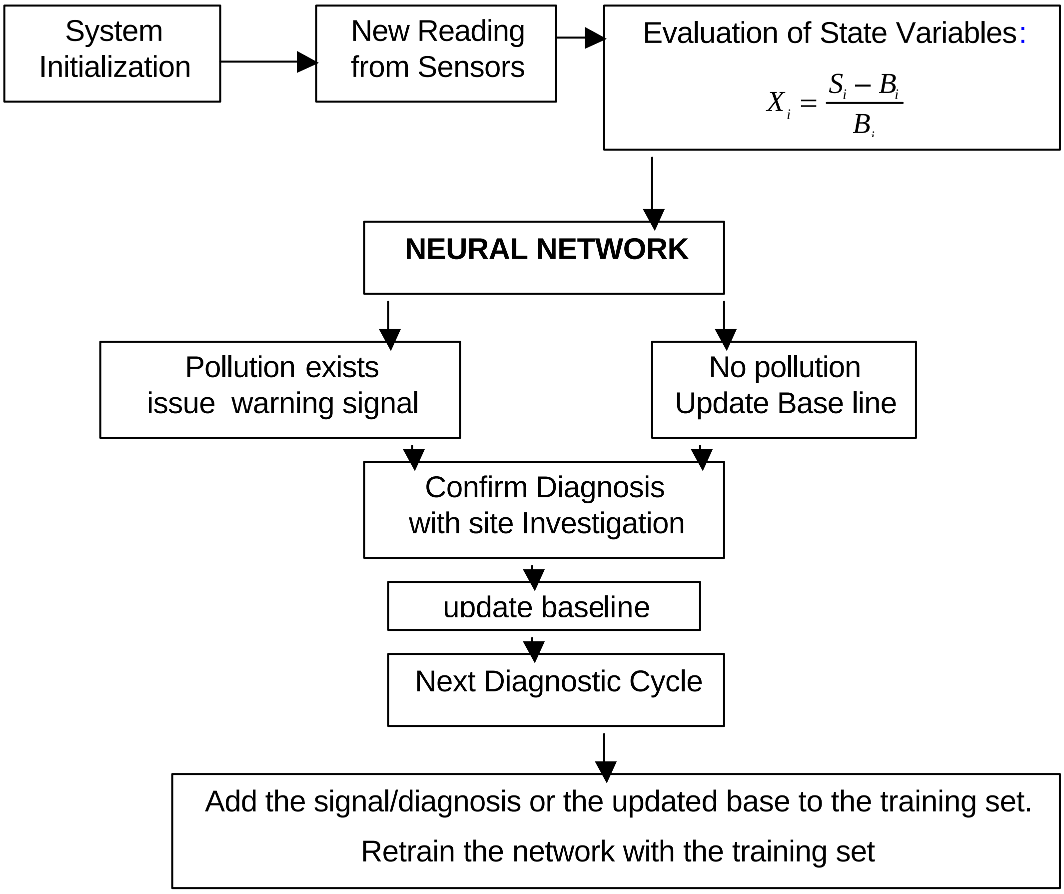


Figure 2 Flowchart of the pollution detection system

The following actions have been taken in order to speed up the learning phase:

• Each channel component of the input vector has been roughly centered and scaled to the sigmoid's range order of magnitude. This ensures that the dot products with small centered initial random weights propagate near zero net inputs to the network units thus avoiding early saturation effects that were experienced in some initial experiments (1).

• I have used the 'on-line' version of the back propagation algorithm. Its updating of the weights after every pattern presentation speeds up learning by avoiding multiple pieces of information collapsing into a single figure for weight updating.

- At the output layer the range of the activation functions has been taken wider than the [-1,1] representation of the target values. It is thus- easier for the network to learn non-asymptotic values. A thresholded version of the quadratic error criterion has been used: no error is back-propagated from an output unit whose current state is >1 (or < -1) when the current target is 1 (or -1).

5. NETWORK ARCHITECTURES AND PERFORMANCE

The major part of the experiments have been carried out with neural networks having two layers of hidden units, following suggestions in [2]. The number of units of the first hidden layer has been varied between 12 and 18, that is two and three times the number of features of the input vector. The second hidden layer was eventually three times as big as the first one.

These architectures involve a high number of parameters with respect to the dimension of the training set, and damaging over-fitting effects might be feared. However, the network classifier with four layers of units proved to be very robust against the growth of the size of the hidden layers. Allowing more unit resources results in over-fitting of the training data, as shown by the top-down increasing discrepancy between the learning and generalization rate figures of table 1. Nevertheless, the corresponding slight improvement of the generalization rate indicates that some rather useful features have been extracted during the overfilling process.

TABLE 1. Classification accuracies for different network architectures

NETWORK ARCHITECTURE	LEARNING RATE (%)	GENERALISATION RATE (%)	OVERALL RATE (%)
6-18-24-20	83.3	77	79.1
6-14-46-20	94.3	77.1	82.8
6-16-52-20	96	77.2	83.5
6-18-54-20	98.2	77.7	84.5

As a direct consequence of these good scaling properties, one can confidently rely upon a correct classification of about 77% of the samples when the network classifier is applied on the total test area. There is also a strong suggestion that the 77% figure would be very close to the best achievable classification rate, for the given site data, when the model class is restricted to fully-connected multilayer perceptrons. However it is planned in the near future to investigate pruning techniques.

The network of the order of 600 epochs was required to reduce the system error to a useful level. This took approximately 2 hours of cpu time on the computer. We hope to produce a parallel transputer implementation of this simulation to reduce the overall training time. A class confusion matrix is shown in figure 2. The performance level achieved is encouraging for data of this type although further work is required on comparisons with traditional methods.

6. VALIDATION OF NETWORK

The correct solutions and those produced by the network may be compared in a qualitative manner or in a quantitative manner using a statistical test. The test systems should be selected in a manner that reduces the likelihood of significant bias in results. Usually, this is done by retaining a random sample of the training patterns, which is a satisfactory method provided the original set of training patterns are representative of the systems likely to be presented to the network. There are many situations where solutions to example systems are not readily available and the task of validation is complicated. One way around this problem is to compare the performance of the neural network with an established alternative method of solving the system. Alternatively, human conjecture could be used to evaluate the network's performance.

7. DISCUSSION

In this paper, neural networks are introduced as a promising management tool that can enhance current automation efforts in the environment industry. Basic neural network architectures are described, and their potential applications in environmental engineering and management discussed. A network application is developed for optimum markup estimation. The purpose of this paper has been to evaluate the neural network approach in the context of operational water quality sensing. Multilayer feed-forward perceptron networks have been used to classify water quality data for water cover mapping. Experiments on a test area using multi-temporal multi-spectral images water quality have demonstrated the potential of neural networks for discriminating large numbers of water cover classes. An average classification accuracy of the order of 84% has been obtained for 20 classes. So far we have established that the multilayer perceptron approach offers a good classification method that competes well with the traditional techniques used in water quality. Indeed by using reasonably large network architectures the method seems to work quite well with large numbers of classes which is where problems are normally encountered with the traditional parametric methods. I believe that neural networks have much potential in water quality sensing and I hope to integrate them into operational applications in the future.

SUMMARY AND CONCLUDING REMARKS

The characteristics of neural networks with respect to possible implementation in construction have been potential applications outlined. Neural networks are introduced as a promising management tool that can enhance automation efforts in the construction problems into patterns is presented, and the variables that govern the design of a neural network for a particular problem are discussed. An example network is developed for optimum markup estimation to demonstrate the use of the present procedure and illustrate the potential benefits of neural networks.

A significant part of inherent construction industry problems lend them-selves to from of analogy-based problems. Some typical examples include everyday decisions on a construction sit, last-minute bid estimating, and design under pressure for speedy resolution. Decisions regarding these problems are made, mainly. On the basis of analogies from past experience, rather than upon detailed analysis of the elements of a given situation. Neural networks, theories and architectures are more suitable in modeling construction industry problems requiring analogy–based solutions then either traditional decision analysis techniques (probabilistic methods, fuzzy sets theory, etc.) or conventional expert systems. Neural networks are less computationally intensive and require limited software to function. They also exhibit superior performance regarding analogy-based decision problems and pattern recognition tasks. This is due to their powerful capabilities for generalizing a solution by learning a few examples, the associative and adaptive processing, and their fault tolerance. Also, their parallel and dis-tributed architecture is well suited to future implementations of parallel processing architectures.

REFERENCES

1. Ayman F. Batisha (1999) “Pollution Detection in Environmental Systems based on Feature-Sensitive Neural Networks” International Conference on Integrated Management of Water Resources in the 21st Century, Cairo, Egypt.
2. Ayman F. Batisha, El-Shinnawy I. A., Nour El-Din M. A.(1999) “Monitoring Thermal Environment based on Neural Networks” International Conference on Integrated Management of Water Resources in the 21st Century, Cairo, Egypt.
3. Ghaboussi, J., and Banan, M. R. (1994). “Neural networks in engineering diagnostics, SAE.” 45th Annu. Earthmoving Industry Conf., Tech. Paper 941116.
4. Goodman. R.M., C.M. Higgins, J. W. Miller, and P. Smyth, 1992. “Rule-based neural networks for Classification and probability estimation.” Neural Computation 4, 781-804.
5. Kanellopoulos I. and Wilkinson G. G. (1990): "Experiments with Backpropagation Neural Networks For Image Classification", Technical Note 1.9064., July 1990, Commission of the European Communities, Joint Research Centre, Ispra, Italy.
6. Pao Y.H. (1989): 'Adaptive Pattern Recognition and Neural Networks', Addison-Weslcy, Reading, Mass.